ADVANCING RESEARCH IN CORRECTIONS
School of Criminology and Criminal Justice | University of Nebraska – Omaha
6001 Dodge Street, Omaha, NE 68182 · 402.554.2610 · arcorrectionslab.org

Alaska Localized LSI-R Violence and Recidivism Assessment

Development and Validation of Localized General and Violent Reconviction Scores for the Alaska Level of Service Inventory-Revised

Prepared for the Alaska Court System, Administrative Office of the Courts

Prepared by: John Ursino, Ph.D. and Zachary Hamilton, Ph.D.

This project was supported by Grant No. 15PBJA-23-GK-02436-DGCT awarded by the Bureau of Justice Assistance, Office of Justice Programs, U.S. Department of Justice. Points of view or opinions in this document are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice

Table of Contents

Executive Summary iii
Primary Findings iii
Purpose and Background 1
The Level of Service Inventory-Revised 2
The Importance of Localization and Revalidation 2
Outcome-Specific Scores 2
Project Structure 6
Validation of the Existing LSI-R 8
Validation Data and Methods 8
Findings 9
Predictive Accuracy 9
Performance Across Subgroups 10
Calibration 11
Implications for Tool Development 12
Localizing the LSI-R 13
Study Population 13
Information Available for Development 14
Reconviction Outcomes 15
Training and Test Samples ..... **Error! Bookmark not defined.**
Assessment Items Considered 15
Criminal History and Prior Offenses 17
Needs and Circumstances 17
Age and Sex 18
Predictive Performance 18
Performance of the Localized LSI-R 18
Fairness Assessment 21
How Fairness Was Evaluated 22
Findings by Sex 23
Findings by Race 25
Comparison of Weighted and Existing LSI-R Score 27
Sources of Improved Predictive Performance 27

Interpreting Distinct Predictive Patterns.....**Error! Bookmark not defined.**
Discussion: Implications for Alaska 30
Using the Scores and Next Steps..... 31
Implementation and Ongoing Evaluation..... 32
Recommended Next Steps 33
Limitations 34
Conclusion..... 36
References..... 38
Technical Appendix 40

Executive Summary

Treatment courts were developed to address substance use and mental health needs that contribute to individuals' involvement with the justice system. Although many early programs excluded individuals with violent offense histories, current treatment court standards emphasize individualized consideration of these applicants when legally permissible (All Rise, 2026). This shift has increased the importance of accurately assessing violent recidivism risk to inform eligibility decisions, supervision practices, and treatment planning. However, many correctional risk assessments were developed to predict general reoffending rather than violence specifically, potentially limiting their usefulness for these decisions.

Alaska has used the Level of Service Inventory-Revised (LSI-R) to assess individuals supervised on probation. This project sought to examine how effectively the LSI-R predicts general and violent reconviction and whether its predictive accuracy can be improved using Alaska's own assessment and outcome data. The study aim was to strengthen the information available to probation agencies and treatment courts while building on the state's existing investment in structured risk assessment.

This work was completed under the Adult Treatment Court Training Initiative, funded by the Bureau of Justice Assistance and awarded to All Rise. The broader initiative seeks to improve the assessment and treatment of individuals with histories of violent offending who may be appropriate candidates for treatment courts. Alaska was one of three states contributing data to develop state-specific assessment tools, with its participation focused on evaluating and refining the LSI-R for use in probation and treatment court decision-making.

Primary Findings

The project first validated the LSI-R as currently scored in Alaska using assessments of the state's probation-supervised population conducted between September 2015 and October 2024 (Ursino & Hamilton, 2026). The primary findings from that evaluation are summarized below.

- *The LSI-R demonstrated weak accuracy for general reconviction and moderate accuracy for felony and violent reconviction.*
- *The LSI-R overpredicted reconviction across outcomes, with predicted risk exceeding observed recidivism rates by roughly 24% for general reconviction and 45% for violent reconviction on average.*

The study findings indicated that the LSI-R's existing scoring structure provides limited predictive accuracy for violence risk and would substantially overstate the likelihood of violent reconviction if applied in treatment court screening. Since the LSI-R items themselves capture relevant risk factors, they were retained and used to develop localized General and Violent Reconviction scores, which were weighted using Alaska assessment and administrative data. Performance was evaluated and the primary findings are summarized below.

- *The localized General and Violent Reconviction tool scores both demonstrated strong predictive accuracy.*
- *Both localized scores were well calibrated, with predicted and observed rates differing by 9% for general reconviction and 6% for violent reconviction on average.*
- *Notably, the two tools draw on different information. The Violent Reconviction tool is driven most strongly by social isolation, mental health problems, and substance use, while the General Reconviction score is driven most strongly by criminal history, substance use, and social isolation.*
- *Predictive accuracy remained strong across sex and racial subgroups. Both tools demonstrated strong predictive performance for men, women, White, and Non-White individuals, with the General Reconviction tool performing somewhat better for women and Non-White individuals.*

Together, the findings indicate that Alaska can obtain substantially improved violence risk prediction with an updated, locally weighted assessment developed without adding items or

changing data collection procedures. Further, because the two tools draw on different risk factors, they allow treatment court teams to distinguish participants whose risk is primarily for general reoffending from those whose risk is specific to violence, rather than treating any history of violence as disqualifying. Overall, the updated tool will provide Alaska agencies with more accurate prediction of general and violent reconviction risk for all the supervised population.

Purpose and Background

Treatment courts integrate judicial supervision with substance use and mental health treatment to address factors contributing to criminal behavior. Many courts have historically excluded individuals with current or prior violent offenses, in part because federal treatment court funding restricted participation by individuals defined as violent offenders (34 U.S.C. § 10611). Exclusions based on charge type alone, however, do not distinguish between individuals with a single violent conviction and individuals whose current circumstances indicate a meaningful likelihood of future violence. Recognizing this limitation, practice standards for treatment courts encourage admission decisions based on validated assessment rather than categorical exclusion (National Association of Drug Court Professionals, 2018).

Actuarial risk assessment provides one way to make these distinctions. Rather than relying solely on the nature of the current charge or on individual discretion, an actuarial assessment examines how case characteristics relate to subsequent outcomes among previously assessed individuals and combines those characteristics into a structured estimate of relative risk. Under the Risk-Need-Responsivity (RNR) framework, the intensity of supervision and services should generally match an individual's assessed risk, while intensive intervention with lower-risk individuals may produce limited benefit, or increase, criminal behavior (Andrews et al., 1990; Bonta & Andrews, 2007; Lowenkamp et al., 2006). As such, an assessment that separates general and violent risk allows treatment court teams to identify participants who can be served safely, those who may require additional supervision and services, from those for whom participation may not be appropriate or responsive to treatment court programming.

The value of risk assessment lies in providing an informed basis for correction, treatment, and case-management decisions. By examining patterns of reoffending among individuals with

similar characteristics, these instruments estimate differences in the likelihood of future offending (Bucklen et al., 2021). Such estimates provide a consistent empirical foundation for professional judgment, allowing treatment court teams to consider assessed risk, alongside other relevant information, when making eligibility, supervision, and treatment decisions. Their usefulness, however, depends on how accurately the underlying instruments assess recidivism risk among the populations they serve. This consideration is particularly relevant in Alaska, which assesses its probation population with an instrument developed in another country and adopted for local use.

The Level of Service Inventory-Revised

The Level of Service Inventory – Revised (LSI-R) was developed in Canada in 1995 using on a sample composed primarily of White, male probationers (Andrews & Bonta, 1995; Vose et al., 2008). The LSI-R combines static criminal history information with dynamic needs that can change over time. In total, the LSI-R contains 54 items across ten subcomponents: Criminal History, Education/Employment, Financial, Family/Marital, Accommodation, Leisure/Recreation, Companions, Alcohol/Drug Problems, Emotional/Personal, and Attitudes/Orientation. Each item contributes one point when the risk factor is present, and the items are summed to produce a total score used to assign a risk level (Andrews & Bonta, 1995). The relative ease of scoring has contributed to its popularity, as the LSI-R is among the most widely used assessments, having been adopted by an estimated 900 correctional agencies across North America and is used in prisons, jails, and community corrections (Vose et al., 2008).

The Importance of Localization and Revalidation

Although the LSI-R remains one of the most widely used risk and needs assessment instruments in correctional settings, its predictive performance has not been consistent across

jurisdictions or over time. Developed using Canadian correctional samples nearly three decades ago, the instrument reflects relationships between risk factors and recidivism observed in the populations on which it was originally constructed and validated. These relationships are not stable across different populations, particularly when jurisdictions differ in their demographic composition, criminal justice policies, sentencing practices, or approaches to community supervision (Hamilton et al., 2022). The result is predictive shrinkage, whereby an assessment demonstrates lower accuracy when applied to populations or settings that differ from those used in its development (Desmarais et al., 2022; Hamilton et al., 2022, 2025).

Assessment accuracy is not fixed over time either, even when the instrument initially performs well in a particular jurisdiction. Tool accuracy can diminish as population characteristics and agency context gradually change over time. Consequently, adopting a new instrument requires evaluating its performance using local data, and periodic revalidation is necessary to determine whether its scores continue to provide accurate and useful information for correctional decision-making (Bucklen et al., 2021).

Research examining the LSI-R across jurisdictions reinforces the importance of evaluating its performance in the populations where it is used. A meta-analysis of 128 studies involving more than 130,000 individuals found that Level of Service instruments demonstrated greater predictive accuracy using samples in Canada, where the tool was developed, than in U.S. samples, suggesting that the instruments' performance may be influenced by the populations and correctional contexts in which they are applied (Olver et al., 2014). Within the U.S., the LSI-R's performance has been found to be inconsistent across populations, assessment practices, and outcomes. These findings suggest that evidence of predictive validity established in one jurisdiction cannot be assumed to apply uniformly elsewhere.

The need for local validation becomes even more important when assessments are used to inform decisions involving different types of reoffending. Although general and violent recidivism sometimes share underlying risk factors, instruments do not necessarily predict these outcomes with the same accuracy. This distinction is particularly consequential for less common outcomes such as violence, where relatively low rates of reoffending make it difficult to identify individuals who are likely to commit a violent offense. Even when an instrument consistently distinguishes individuals at greater risk of violent reoffending from those at lower risk, it may still overestimate the likelihood that violence will occur (Steyerberg, 2019). Consistent with these concerns, the Level of Service instruments generally predict violent recidivism less accurately than general recidivism (Olver et al., 2014) and its ability to predict violent reoffending is inconsistent across samples (Campbell et al., 2009).

Limitations in outcome-specific prediction have direct implications for how the LSI-R is used by local and state agencies, particularly when assessment results inform treatment or program eligibility and other decisions involving public safety. The instrument was designed primarily to assess general risk and criminogenic needs, rather than to produce separate estimates of violent reconviction. Under its existing scoring structure, each item contributes one point to the total score. This approach does not account for differences in how strongly individual items predict recidivism. For example, an item strongly associated with subsequent offending contributes no more to the total score than one with a relatively weak relationship to that outcome.

The predictive importance of individual items may also differ depending on the outcome being assessed, where an item may be particularly informative for predicting general recidivism while being less useful for predicting violent offenses, and vice versa. Consequently, the existing

scoring structure may not fully capture the information contained in the LSI-R items, particularly when used to inform decisions about violence risk. For treatment courts, accurately estimating violent risk is especially important, as overestimation may unnecessarily restrict access to treatment-oriented alternatives, while underestimation may raise legitimate public safety concerns.

Local revalidation and refinement of the existing scoring structure offer a practical means of addressing these limitations while preserving Alaska's existing assessment infrastructure. Research suggests that the predictive contributions of individual risk factors can vary across jurisdictions, populations, and supervision contexts, supporting the use of local outcome data to evaluate and refine established instruments (Desmarais et al., 2022; Hamilton et al., 2022). Rather than assuming that each LSI-R item should contribute one point to the total score, an empirically derived scoring structure can assign weights based on observed relationships with recidivism among individuals supervised in Alaska. The same information can also be used to develop outcome-specific scoring models, allowing existing assessment items to generate distinct estimates of general and violent recidivism risk. This approach builds on the state's existing investment in structured assessment while improving the extent to which resulting scores reflect its current probation population and the particular outcomes relevant to supervision and treatment court decisions.

Outcome-Specific Scores

The development of outcome-specific risk scores reflects a broader recognition that recidivism is not a single, uniform outcome. Individuals who are more likely to commit a new offense are not necessarily those most likely to commit a violent offense, and the factors associated with each outcome may differ in their relative importance. A single general risk score

may therefore obscure meaningful differences in the types of reoffending individuals are most likely to engage in. Developing separate scores allows the same assessment information to be weighted according to its relationship with each outcome, producing more targeted estimates of risk without requiring additional information collection. This distinction is particularly relevant when assessment results inform supervision intensity decisions, treatment placement, or eligibility for programs where the likelihood of violent offending is an important consideration.

The current project applies these principles by developing separate General Reconviction and Violent Reconviction scores, each with its own empirically derived item weights. These scores are intended to answer related, but distinct, questions rather than represent different levels of severity along a single risk continuum. The project also recognizes that jurisdictions differ in the assessment information available to agencies and their capacity to develop locally tailored instruments. Accordingly, the work includes both the refinement of Alaska's existing LSI-R instrument and the development of pooled assessment models intended for use in settings where local model development is not feasible. The following section describes how these components were organized within the broader project.

Project Structure

This report is one product of a multistate effort to develop violence risk tools for treatment court populations. The original design proposed a single unified violence assessment developed from pooled data across participating states. Because the participating states use different assessment systems, with Alaska and South Dakota using the LSI-R and Arizona using the Offender Screening Tool (OST), there was not enough overlap in available items to responsibly build one common instrument. The project was therefore revised to produce:

1. *A state-specific, localized assessment built from each participating state's own instrument and outcome data; and*

2. *Two pooled violence assessments, one for jurisdictions using the LSI-R and one for jurisdictions using the Ohio Risk Assessment System (ORAS), for courts that cannot develop a jurisdiction-specific tool.*

The present report documents the Alaska state-specific assessment. The pooled assessments will be described in a separate report. The next section presents the validation of the existing LSI-R. The remainder of this report then describes the population and outcomes used to develop the assessment, the items considered for inclusion, the methods used to construct and validate the General and Violent Reconviction tools, performance across demographic groups, and the translation of the statistical models into a scoring system suitable for agency use.

Validation of the Existing LSI-R

Before developing the localized assessment, the research team validated the LSI-R as currently scored in Alaska (Ursino & Hamilton, 2026). The validation assessed how accurately the LSI-R predicted reconviction within three years of assessment, whether accuracy differed by sex and race, and whether predicted risk aligned with observed reconviction rates. Its findings establish the baseline against which the localized assessment should be judged. The analyses also identify the aspects of existing performance most relevant to outcome-specific redevelopment.

Validation Data and Methods

The validation used LSI-R assessment records from Alaska's probation-supervised population, drawn from 1,953 assessments conducted between September 2015 and October 2024. Predictive performance was evaluated for three outcomes: general (or any) new reconviction, felony reconviction, and violent reconviction, occurring within three years of assessment. For each outcome, performance was examined along two complementary dimensions: discrimination and calibration.

Discrimination measures how well an instrument distinguishes individuals who are subsequently reconvicted from those who are not. An instrument with stronger discrimination generally assigns higher risk scores to individuals who are reconvicted. Discrimination was measured using the Area Under the Receiver Operating Characteristic Curve (AUC), where values below 0.56 indicate negligible accuracy, 0.56 to 0.63 weak accuracy, 0.64 to 0.70 moderate accuracy, and 0.71 or greater strong accuracy (Rice & Harris, 2005). National guidelines identify an AUC of 0.64 as the minimum threshold for acceptable performance in criminal justice risk assessment (Desmarais et al., 2022).

Calibration examines whether estimated probabilities of reconviction correspond to actual outcomes. For example, among individuals assigned a 20% probability of reconviction, approximately 20% should be reconvicted if the estimates are accurate. Calibration was measured using the Expected Calibration Error (ECE), which summarizes the average difference between predicted and observed reconviction rates across score levels. An ECE of 0.03 indicates that predicted risk differed from observed reconviction rates by approximately 3%, on average. The distinction between discrimination and calibration is important for interpreting the validation results. An instrument may successfully identify individuals at relatively higher risk while still over-, or under-, estimating their actual likelihood of reconviction.

Findings

The following section describes findings from study analyses. As described, we examined the overall predictive performance of the LSI-R, but also whether that performance varies across demographic subgroups and whether estimated risk corresponds to observed outcomes. Results are presented first for discrimination, followed by subgroup differences and calibration.

Predictive Accuracy

Table 1 displays AUCs for the LSI-R across the three outcomes. The LSI-R demonstrated weak accuracy for general reconviction (AUC = 0.63), falling just below the threshold for acceptable performance, and moderate accuracy for felony reconviction (AUC = 0.65) and violent reconviction (AUC = 0.66). Thus, the LSI-R performed somewhat better for more serious forms of reoffending, but its overall predictive accuracy remained modest. Out of the 15 tests conducted, weak predictive discrimination was identified 3 times and a moderate effect 12 times, and no test reached the strong range.

Table 1. AUCs for the LSI-R Across Reconviction Outcomes

Group	General	Felony	Violent
Total	0.63	0.65	0.66
Male	0.64	0.65	0.66
Female	0.61	0.66	0.65
White	0.64	0.67	0.66
Non-White	0.62	0.64	0.65

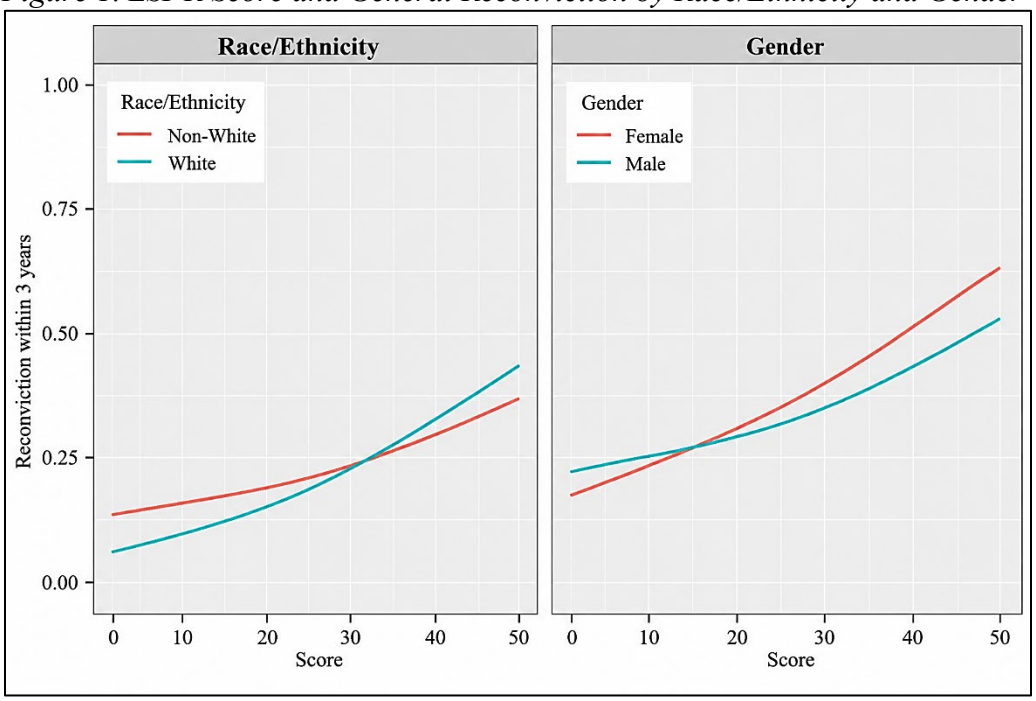
Note: AUC values below 0.56 indicate negligible accuracy, 0.56 to 0.63 weak accuracy, 0.64 to 0.70 moderate accuracy, and 0.71 or greater strong accuracy.

Performance Across Subgroups

Differences by sex were modest. Accuracy was slightly lower for women than for men for general reconviction (AUCs of 0.61 and 0.64) but comparable for felony reconviction (AUCs of 0.66 and 0.65) and violent reconviction (AUCs of 0.65 and 0.66). Differences by race were also modest but followed a consistent pattern, with accuracy higher for White than for Non-White individuals across all three outcomes. The largest gap appeared for felony reconviction, where accuracy was well within the moderate range for White individuals (AUC = 0.67) and at the lower boundary of that range for Non-White individuals (AUC = 0.64).

Figure 1 illustrates this pattern for general reconviction. For both groups, the predicted probability of reconviction increases as LSI-R scores rise. Non-White individuals show a higher predicted probability near the lower end of the score range, with approximately a 15% difference between White and Non-White individuals near the low-end of the scale. The lines converge and White individuals have about a 10% higher probability of reoffending on the high-end of the scale. The steep increase in probability for White individuals indicate that increases in risk score are more strongly associated with increased recidivism risk for White individuals that for Non-White individuals. The same pattern occurs for gender, where initial probabilities are higher for males, but the lines cross and separate as scores increase.

Figure 1. LSI-R Score and General Reconviction by Race/Ethnicity and Gender



Calibration

Table 2 displays ECEs for the LSI-R. Calibration errors were smallest for general reconviction, with predicted and observed rates differing by roughly 24 percentage points. Errors were considerably larger for felony and violent reconviction, with predicted risk exceeding observed rates by about 45 percentage points for each. Errors of this size, particularly the overprediction of violent reconviction, may lead to inappropriate supervision and treatment decisions with tool scores providing inflated estimates of risk.

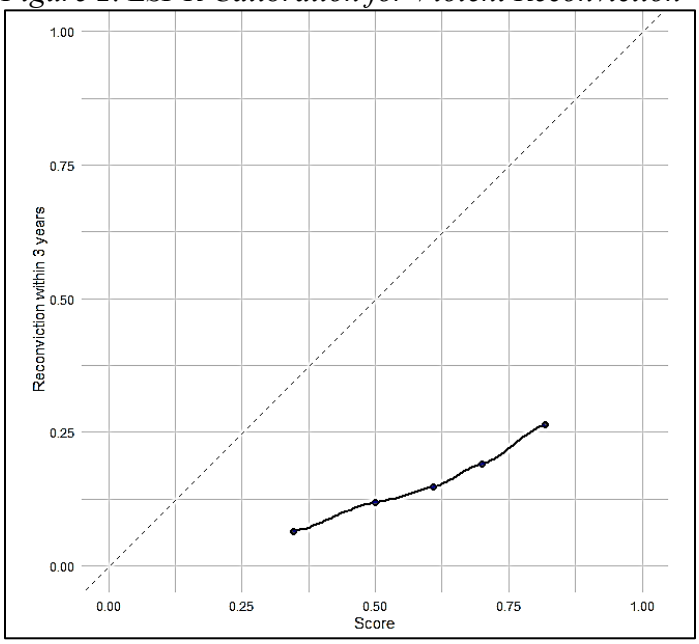
Table 2. ECEs for the LSI-R Across Outcomes

Outcome	ECE
General Reconviction	0.24
Felony Reconviction	0.45
Violent Reconviction	0.45

Figure 2 presents the calibration results for violent reconviction. The diagonal dotted line represents perfect calibration, where predicted probabilities match observed reconviction rates.

The calibration trend remains well below the diagonal across the observed score range, indicating that the LSI-R substantially overpredicts violent reconviction. The gap is most pronounced at higher predicted scores, where individuals with a predicted probability near 0.80 had an observed violent reconviction rate closer to 0.25. Consequently, individuals may be assigned to higher risk categories than their observed likelihood of violent reconviction would warrant, affecting supervision intensity and treatment decisions.

Figure 2. LSI-R Calibration for Violent Reconviction



Implications for Tool Development

The evaluation found that the LSI-R's predicted probabilities substantially overstated reconviction risk, particularly for felony and violent outcomes. The findings reinforce the importance of developing outcome-specific scoring models and informed the selection of assessment data for the localized models. A localized tool reweights the LSI-R's current assessment items and adds useful predictors from regularly collected administrative information. Item weighting is completed separately for general and violent reconviction, allowing each factor's contribution to reflect its relationship with the outcome being predicted rather than

relying on the existing total score. The sections that follow describe the development of these models and evaluate whether they improve on the performance documented here.

Localizing the LSI-R

Because the initial validation established the need for outcome-specific scoring, the next phase used Alaska assessment and administrative records to develop separate General and Violent Reconviction models, then evaluated both models on data withheld from development. This cross-validation approach allows the development results to be assessed against a previously documented baseline rather than interpreted in isolation. The following sections describe the sample, predictors, and estimation procedures.

Study Population

The localized LSI-R was developed using records for 1,953 individuals assessed with the LSI-R in Alaska between September 2015 and October 2024. Records were drawn from LSI-R assessments of Alaska's probation-supervised population. The records allowed the research team to observe each individual's assessment information and then determine whether a new conviction occurred during the three years that followed. This structure maintains a clear separation between the information available when a decision is made and the outcome that occurs afterward.

To develop and evaluate the localized OST, the study population was randomly divided into a 70% training sample and a 30% test sample, stratified by reconviction outcome to ensure similar proportions of individuals who were reconvicted in each sample. The training sample was used to identify the assessment items associated with reconviction and estimate their respective scoring weights. These weights were then applied to the independent test sample to evaluate how accurately the completed assessment predicted reconviction among individuals who were not

involved in its development. The test sample was not used to select items or estimate weights, providing a more realistic estimate of predictive performance than evaluating the assessment using the same individuals on which it was developed. Table 3 summarizes the study population, with the training and test samples displayed separately to document the comparability of their characteristics.

Table 3. Characteristics of the Alaska Study Population

Characteristic	Training Sample	Test Sample	Total
Individuals (N)	1,367	586	1,953
Male (%)	69.4	72.1	70.2
White (%)	47.3	50.0	48.1
Non-White (%)	52.7	50.0	51.9
General Reconviction within 3 years (%)	37.1	37.2	37.1
Violent Reconviction within 3 years (%)	17.1	16.4	16.9

Information Available for Development

The project did not begin with a new list of risk factors. Instead, we evaluated information Alaska already collects through the LSI-R and its supervision records. Candidate measures came from three existing sources:

- *LSI-R assessment items covering criminal history, education and employment, finances, family and marital relationships, accommodation, leisure and recreation, companions, alcohol and drug problems, emotional and personal functioning, and attitudes.*
- *Criminal history information, including counts of prior felony offenses, prior violent offenses, and total prior offenses.*
- *Demographic information, including age and sex.*

Most LSI-R items record whether a risk factor is present or absent. Thirteen items are rated by the assessor on a 0 to 3 scale, where lower ratings indicate greater need. Consistent with standard LSI-R scoring, ratings of 0 or 1 were treated as indicating that the risk factor was present and ratings of 2 or 3 as indicating that it was absent. Age was grouped into six

categories: 20 or younger, 21 to 30, 31 to 40, 41 to 50, 51 to 60, and 61 or older. Cases missing an outcome were excluded.

Reconviction Outcomes

An assessment is understood in relation to the outcome it was developed to predict. Two outcomes were selected for this project, both measured over the three years following assessment. A fixed follow-up period allows the two outcomes to be evaluated on a consistent time horizon. The same General and Violent Reconviction outcomes used to evaluate the LSI-R's initial performance were used to develop the new tools. Among individuals with complete three-year follow-up, 37.1% experienced General Reconviction and 16.9% experienced Violent Reconviction.

Assessment Items Considered

Fifty-nine candidate measures were evaluated for each outcome, including 54 LSI-R items and five measures of criminal history and demographic characteristics. Separate models were developed for General and Violent Reconviction to identify which items contributed to predicting each outcome and how much weight they should receive.

Model development accounted for overlap among assessment items. For example, an individual with three or more prior sentencings necessarily also has two or more and is likely to have a prior incarceration and a higher count of prior offenses. Although each measure may be associated with reconviction, they often capture similar information. We used penalized regression to account for these relationships, select the most informative items, and estimate their contributions to each outcome. All items were coded so that higher values indicated greater risk, and items receiving negative weights were excluded from the final models. Additional details on

model development and estimation procedures are provided in Appendices A1 and A2. Table 4 summarizes the candidate measures by subcomponent and the number retained in each model.

Table 4. Candidate and Retained Items by Subcomponent

Subcomponent	Candidate Items	Retained General	Retained Violent	Content
Criminal History	10	8	4	Prior convictions and sentencings, present offenses, early arrest, incarceration, escape, institutional misconduct, supervision violations, record of violence
Education / Employment	10	6	9	Current and past employment, job loss, educational attainment, school suspension, interactions at school or work
Financial	2	0	1	Financial problems and reliance on social assistance
Family / Marital	4	1	1	Marital, parental, and other family relationships; criminal family or spouse
Accommodation	3	3	2	Satisfaction with accommodation, address changes, high-crime neighborhood
Leisure / Recreation	2	1	2	Participation in organized activities and use of leisure time
Companions	5	2	2	Social isolation and criminal and anti-criminal acquaintances and friends
Alcohol / Drug Problems	9	8	6	Lifetime and current alcohol and drug problems and their effects on law violations, family, school or work, and health
Emotional / Personal	5	5	5	Mental health interference, treatment history, psychological assessment
Attitudes / Orientation	4	2	1	Attitudes toward crime, convention, and supervision
Age, Prior Offense Counts, and Sex	5	4	4	Age; prior felony, violent, and total offenses; sex
Total	**59**	**40**	**37**	

The complete item wording and response options are provided in the accompanying scoring workbook. Retention counts are useful for describing model composition, but they

should not be read as a ranking of the causal importance of subcomponents. Items with similar content may cancel out each other's predictive weight, and a subcomponent with many retained items may contribute less to the final score than a subcomponent with only one or two highly weighted items. Likewise, an item assigned no points in one model is not necessarily unrelated to that outcome in isolation but may offer little additional information once other items are considered.

Criminal History and Prior Offenses

Criminal history is consistently among the strongest predictors of future offending (Gendreau et al., 1996). The localized LSI-R draws on this information in two forms: the LSI-R Criminal History items, which record the extent, timing, and seriousness of prior offending and prior responses to supervision and incarceration; and counts of prior felony, violent, and total offenses. These measures' content overlap, and the modeling approach described in the next section outlines the predictive value each provides.

Needs and Circumstances

The Education/Employment, Financial, Family/Marital, Accommodation, Leisure/Recreation, Companions, Alcohol/Drug Problems, Emotional/Personal, and Attitudes/Orientation subcomponents describe an individual's current circumstances and needs. Unlike prior convictions, many of these items can change through intervention or changing life circumstances. Their inclusion means that an individual's scores can change at reassessment when circumstances such as employment, substance use, relationships, or residential stability change. Their predictive contributions should not be taken as evidence that addressing any particular need will necessarily reduce reconviction.

Age and Sex

Age is among the most consistently observed predictors of offending, with risk generally declining across adulthood. Sex was included as a candidate predictor so that the model could account for baseline differences in the rates of reconviction between men and women.

Predictive Performance

Once the General and Violent Reconviction models were finalized using the training sample, they were applied to the test sample. Performance was evaluated using discrimination and calibration, with results compared against the existing LSI-R to determine whether the updated models improved predictive accuracy. Discrimination was measured using the AUC statistic to assess how well each model distinguished between individuals who were and were not reconvicted (Rice & Harris, 2005). Calibration was evaluated by comparing predicted probabilities with observed reconviction rates across ten groups of approximately equal size and summarized using the ECE, with lower values indicating better agreement. Calibration plots were also examined to identify patterns of overprediction and underprediction, and bootstrap confidence intervals were calculated for ECE estimates.

Performance of the Localized LSI-R

Table 5 summarizes discrimination and calibration for the localized version of Alaska's tool using test sample. General Reconviction achieved an AUC of 0.79, and Violent Reconviction achieved an AUC of 0.77, both within the strong range for predictive accuracy. The corresponding ECEs were 0.09 and 0.06, indicating that predicted and observed reconviction rates differed by about 9 and 6 percentage points on average.

Table 5. Test Sample Performance

Outcome	AUC	ECE
General Reconviction	0.79	0.09
Violent Reconviction	0.77	0.06

Although the models demonstrated strong discrimination, their ability to distinguish individuals by risk does not necessarily mean that predicted probabilities accurately reflect the likelihood of reconviction. Accurate probability estimates are particularly important for violent reconviction since violent offenses are relatively uncommon, meaning that individuals identified as high-risk relative to others may still have a relatively low overall probability of being reconvicted for a violent offense. The calibration analyses therefore examine how closely predicted probabilities derived from risk scores correspond to observed reconviction rates.

Figure 3 presents the calibration results for General Reconviction, comparing average predicted probabilities with observed reconviction rates across ten risk groups. Predictions generally tracked observed outcomes across the distribution, although agreement varied more from group to group than it would in a larger sample because each risk group contained about 30 individuals. The largest departure occurred in the seventh risk group, where predicted risk averaged 49.0% and the observed rate was 36.7%, while the model underpredicted in the third risk group (15.5% predicted and 26.7% observed). Agreement was close at the high end of the range, where the two highest risk groups had predicted rates of 72.8% and 81.5% and observed rates of 79.3% and 82.8%. Overall, the updated model provides more accurate estimates of general reconviction risk than the existing LSI-R, which overstated reconviction by 24% on average.

Figure 3. Observed and Predicted General Reconviction

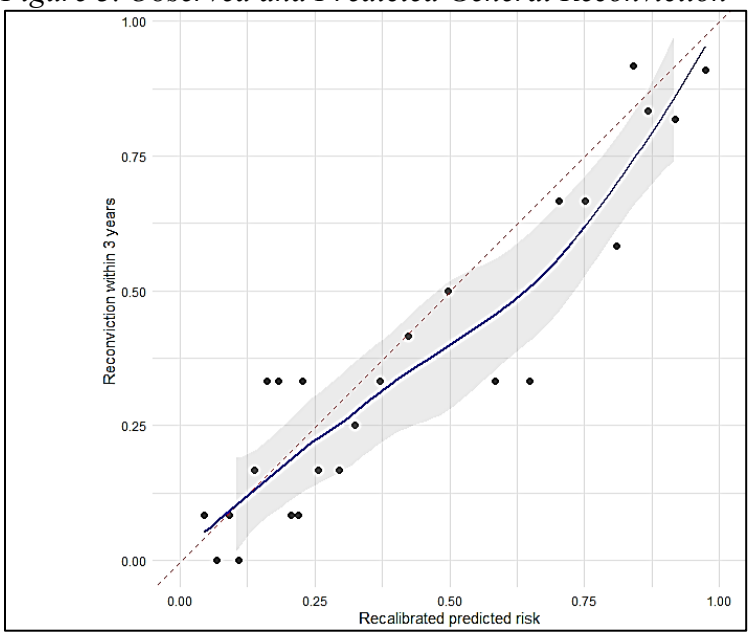
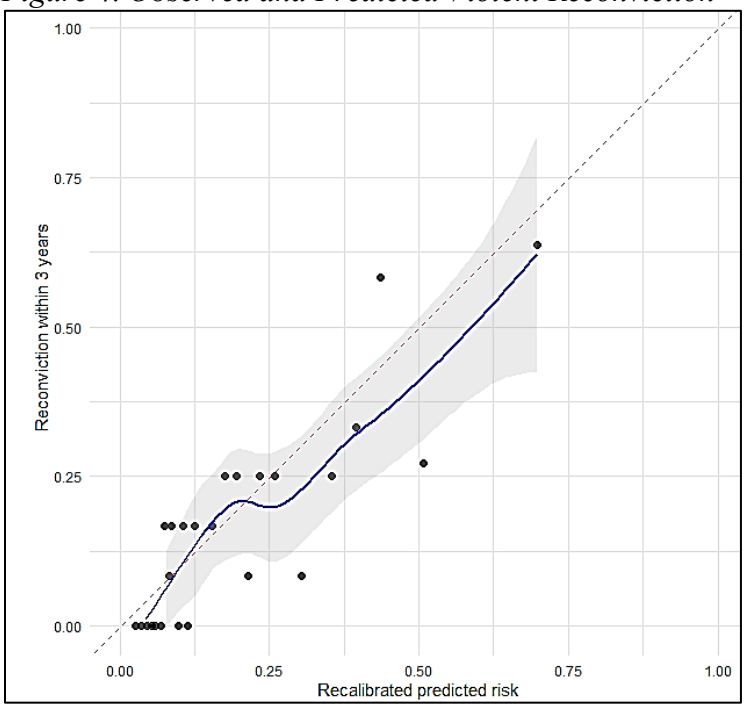


Figure 4 presents the corresponding results for Violent Reconviction. Despite this outcome's lower base rate, predicted probabilities closely aligned with observed rates across most risk groups. The largest departure occurred near the middle of the range, where the model predicted a 13.4% reconviction rate compared with an observed rate of 23.3%. Notably, predicted and observed rates were nearly identical at the highest-risk end of the scale (48.1% predicted and 48.3% observed). Taken together, the updated model provides more accurate estimates of violent reconviction than the existing LSI-R, which overstated violent reconviction by about 45%, on average.

Figure 4. Observed and Predicted Violent Reconviction



Fairness Assessment

An assessment can perform well for the overall population while performing differently for particular groups. For that reason, the localized LSI-R was also evaluated separately by sex and by race. The purpose of this analysis was not simply to determine whether groups received different average scores. Groups can differ in their observed rates of reconviction, and score distributions may differ even when an assessment predicts equally well. The analysis instead examined whether the assessment functioned similarly when predicting reconviction within each group.

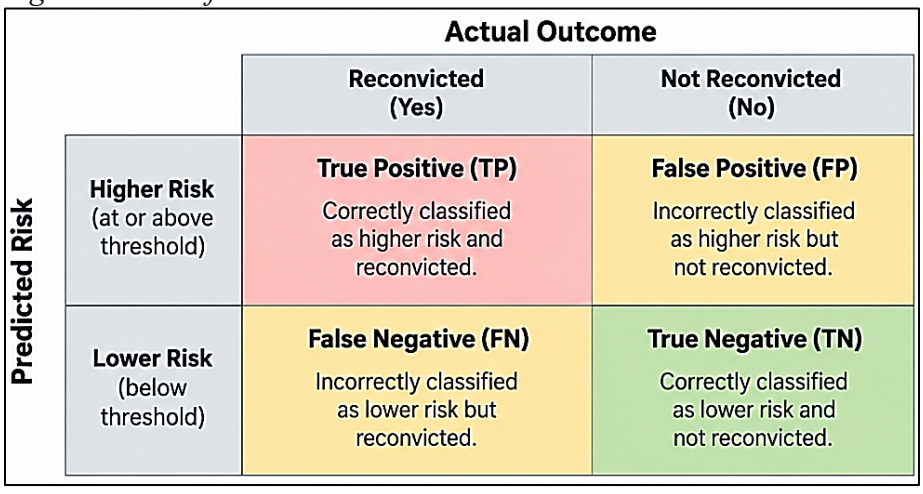
As stated, race was not used to calculate either score and was retained only to evaluate tool fairness. Sex was included as a candidate predictor so that each model could account for differences in reconviction rates between men and women, although its estimated contribution was small enough that it receives no points on the scorecard. These choices should be considered

when interpreting both the subgroup comparisons and the operational implications of the scores. The exclusion of race does not rule out group differences arising through other predictors.

How Fairness Was Evaluated

Predictive performance was examined across demographic groups using AUCs and comparisons of predicted and observed reconviction rates. Classification errors were also examined by applying the same risk threshold across groups for each outcome. Figure 5 illustrates the four possible classification outcomes. True positives and true negatives represent correct classifications, while false positives and false negatives represent incorrect classifications. The false-positive rate (FPR) measures the percentage of individuals who were not reconvicted but were classified as higher risk. The false-negative rate (FNR) measures the percentage of individuals who were reconvicted but were classified as lower risk. Comparing these rates across demographic groups helps identify whether certain groups are more likely to experience either type of classification error.

Figure 5. Classification Outcomes and Error Rates



To compare classification errors across groups, we used reference thresholds of 33% predicted probability for General Reconviction and 11% for Violent Reconviction. These

thresholds were used solely for comparison and are not recommended for classification or eligibility decisions. Appendix A4 presents alternative thresholds and their associated error rates. For the racial subgroup analyses, individuals were classified as White or Non-White based on available administrative records. The Non-White category combined multiple racial and ethnic groups because the number of reconviction events within individual groups, particularly violent reconvictions, was insufficient to produce reliable estimates. Although this aggregation improved statistical stability, it does not imply that the groups have similar outcomes and limits the ability to identify differences among them.

Findings by Sex

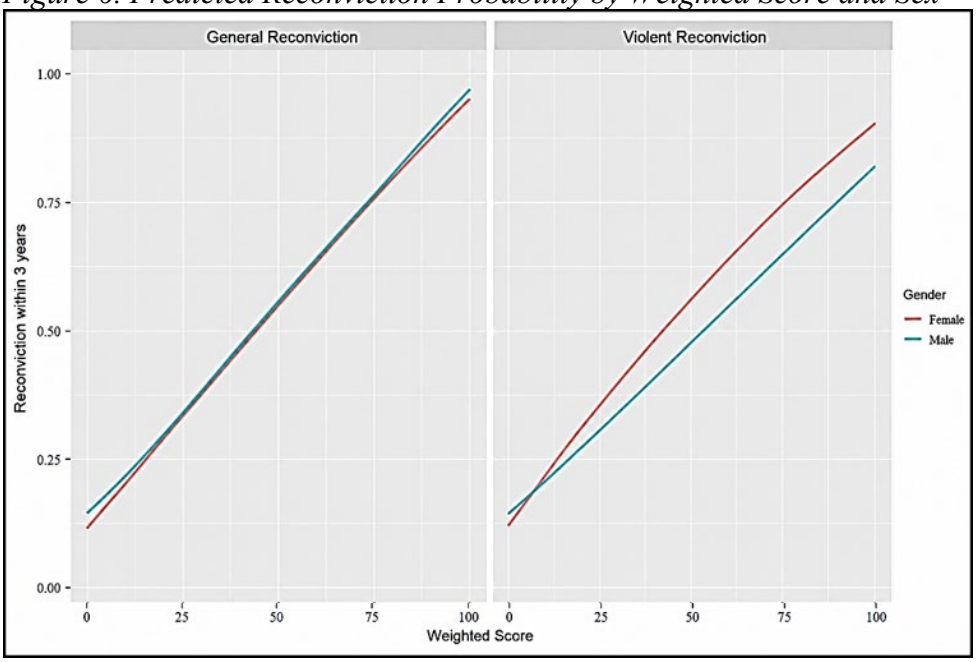
Predictive accuracy was strong for both men and women, although both models performed better for women. The General Reconviction score demonstrated strong accuracy for men (AUC = 0.78) and women (AUC = 0.87), while the Violent Reconviction score demonstrated strong accuracy for both groups (AUC = 0.77 for men and 0.81 for women). The smaller number of women in the test sample makes the estimates for women less precise. Classification error rates at the reference thresholds were broadly similar for men and women. For General Reconviction, men had a higher false-positive rate than women (28.6% versus 20.4%) and a similar false-negative rate (23.9% versus 22.5%). For Violent Reconviction, false-positive rates (42.5% versus 39.7%) and false-negative rates (14.7% versus 13.3%) were similar for men and women.

Table 6. Performance by Sex, Test Sample

Outcome	Group	Reconviction (%)	AUC	FPR	FNR
General	Men	34.8	0.78	28.6	23.9
General	Women	42.6	0.81	20.4	22.5
Violent	Men	15.8	0.77	42.5	14.7
Violent	Women	18.1	0.80	39.7	13.3

Figure 6 illustrates how predicted reconviction probabilities change across weighted risk scores for men and women. For general reconviction, there is a steep increase in predicted reconvictions as risk score increases for both genders. Moreover, both lines are nearly identical, indicating that increasing risk scores are associated with equitable increases in predicted reconviction risk between genders. The largest gap occurs at the low-end of the scale, with a less than 5% difference. For violent reconviction, we see the same steep increase. However, there is more separation between the groups as score increases. The gap is largest at the high-end of the scale, with a modest difference of approximately 10%.

Figure 6. Predicted Reconviction Probability by Weighted Score and Sex



Taken together, the findings indicate that both models perform well for men and women, with all four estimates in the strong range and somewhat stronger accuracy for women. Overall, the results provide encouraging evidence of comparable predictive performance for men and women, while identifying classification errors as an area warranting continued attention when selecting thresholds for translating tool scores into operational decisions.

Findings by Race

Differences in predictive accuracy by race were modest. The General Reconviction tool demonstrated strong accuracy for White (AUC = 0.77) and Non-White (AUC = 0.80) individuals, and the Violent Reconviction tool also demonstrated strong accuracy for both White (AUC = 0.78) and Non-White (AUC = 0.76) individuals. Accuracy for general reconviction was higher for Non-White individuals, reversing the direction of the gap observed for the existing LSI-R in the validation, while accuracy for violent reconviction was slightly higher for White individuals. Overall, the localized tools demonstrate consistently strong discrimination across groups. However, future evaluations should further monitor group differences to ensure fairness remains.

Classification error rates at the reference thresholds differed across groups. Observed reconviction rates were higher for Non-White individuals for both outcomes (43.2% versus 33.1% for General Reconviction and 20.4% versus 13.4% for Violent Reconviction). For General Reconviction, Non-White individuals had a higher false-positive rate than White individuals (29.8% versus 23.6%) but a substantially lower false-negative rate (18.8% versus 31.8%). For Violent Reconviction, Non-White individuals again had a higher false-positive rate (46.0% versus 38.2%) and a slightly lower false-negative rate (13.8% versus 15.8%).

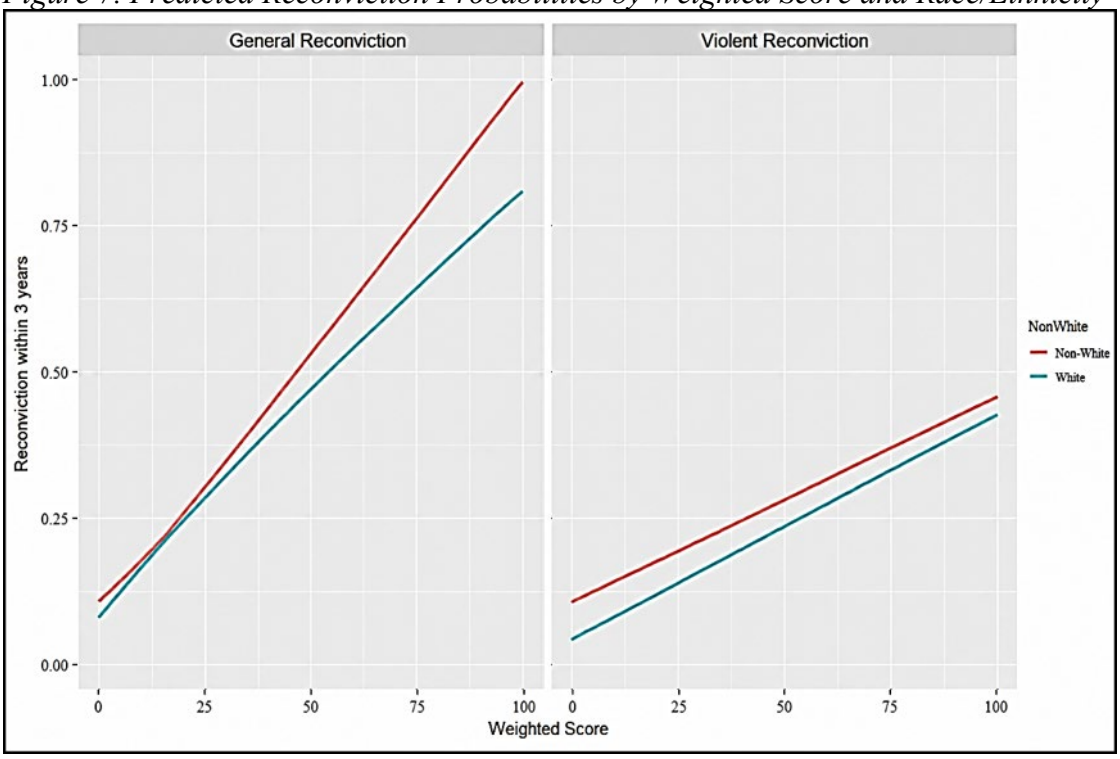
Table 7. Performance by Race, Test Sample

Outcome	Group	Reconviction (%)	AUC	FPR	FNR
General	White	33.1	0.77	23.6	31.8
General	Non-White	43.2	0.80	29.8	18.8
Violent	White	13.4	0.78	38.2	15.8
Violent	Non-White	20.4	0.76	46.0	13.8

Figure 7 illustrates how predicted reconviction probabilities change across weighted risk scores for White and Non-White individuals. For general reconviction, there is a steep increase in predicted reconvictions as risk score increases for both groups. The lines are nearly identical at

the low-end of the scale, but they separate as score increases, with predicted reconviction risk rising faster for Non-White individuals. The gap is largest at the high-end of the scale, with a difference of approximately 15%. For violent reconviction, the increase is much more gradual, and predicted reconviction stays below 50% for both groups, reflecting the relative rarity of this outcome. The lines run nearly parallel across the scale, indicating that increasing risk scores are associated with similar increases in predicted reconviction risk between groups. The largest gap occurs at the low-end of the scale, with a difference of approximately 6%, and it narrows to about 3% at the high-end.

Figure 7. Predicted Reconviction Probabilities by Weighted Score and Race/Ethnicity



Overall, the models demonstrated comparable ability to distinguish individuals by risk across racial categories. Differences in classification errors at a common threshold can be more pronounced than differences in discrimination, which highlights the importance of examining

classification performance across groups rather than relying solely on overall predictive accuracy when evaluating the instruments.

Comparison of Updated and Existing LSI-R Score

The existing LSI-R total score was also evaluated using the same test sample and reconviction outcomes to determine whether the new models predict general and violent reconviction more accurately than the existing assessment.

The earlier validation (Tables 1 and 2) used all available assessments. For a direct comparison, Table 8 evaluates the existing and localized LSI-R. The localized tool demonstrates substantial improvement in predictive discrimination, increase the AUC from a weak effect size (0.63) to a strong effect size (0.79). using the same test sample. Discrimination for violent reconviction similarly increases from moderate (AUC = 0.66) to strong (AUC = 0.77) Calibration error also decreased by 15 and 39 percentage points on average for general and violent reconviction respectively. Improvements in calibration are particularly important for violent reconviction, as the existing LSI-R overstated violent reconviction by about 45 percentage points on average in the validation. Continued use of the existing scoring would therefore substantially overestimate individuals' likelihood of violent reoffending, potentially resulting in unnecessarily restrictive supervision and treatment decisions.

Table 8. Comparison of Localized and Existing LSI-R Scores

Outcome	Original AUC	Localized AUC	Original ECE	Localized ECE
General Reconviction	0.63	0.79	0.24	0.09
Violent Reconviction	0.66	0.77	0.45	0.06

Sources of Improved Predictive Performance

The final models show which items are the strongest predictors for both outcomes. Age contributes substantially to both scores, adding up to 11 General and 13 Violent points for the

youngest individuals, but the remaining measures differ by outcome. These patterns provide context for understanding why separate outcome-specific models were retained.

For General Reconviction, criminal history carries the most weight as a group, with eight of the ten Criminal History items retained, led by two or more prior sentencings and arrest before age 16. Substance use also contributes heavily, particularly substance use that has caused medical problems or contributed to law violations, as do social isolation and prior mental health treatment. These contributions combine information about criminal history with measures of current circumstances. Their weights describe predictive associations after accounting for other retained items.

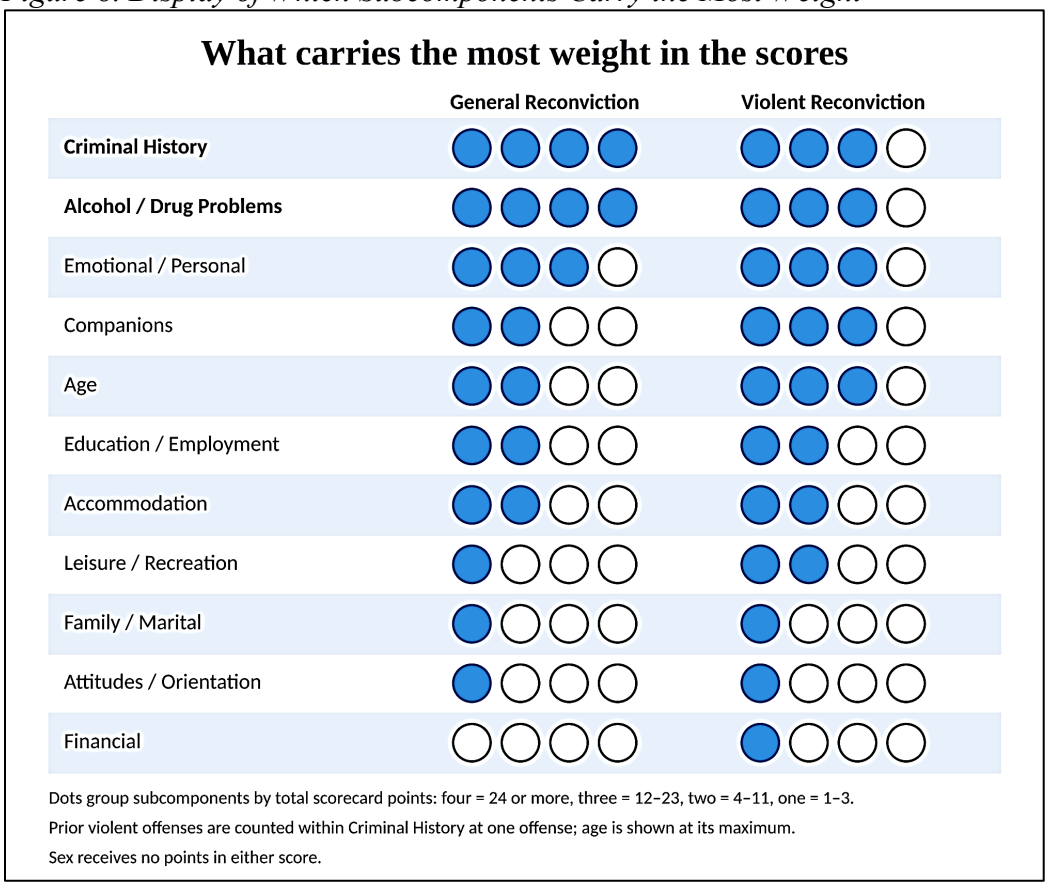
For Violent Reconviction, social isolation is the single largest item, followed by severe interference from mental health problems. Emotional and personal items contribute more to violence than to general reconviction, and nine of the ten Education/Employment items carry weight in the violent model. Criminal history contributes through prior incarceration, three or more prior sentencings, arrest before age 16, and the count of prior violent offenses. The official record of assault or violence carries no Violent points, which likely reflects its overlap with the count of prior violent offenses.

Retaining separate models for general and violent reconviction creates a multiband assessment that provides distinct risk estimates for different outcomes. Prior research supports this approach, demonstrating the value of assessing multiple recidivism outcomes rather than relying on a single overall risk score (Duwe, 2014; 2024; Hamilton et al., 2022; 2025).

The differing contributions of assessment subcomponents highlight the value of developing separate scoring models for general and violent reconviction rather than relying on a single measure of risk. Figure 8 illustrates the relative contribution of each subcomponent to the

General and Violent Reconviction scores, grouped by the total scorecard points each subcomponent carries. Criminal history and alcohol and drug problems carry the most weight in the General Reconviction score, together accounting for about half of the available points. The Violent Reconviction score distributes weight more evenly, with emotional and personal functioning carrying the largest share, followed by criminal history, companions, and alcohol and drug problems.

Figure 8. Display of Which Subcomponents Carry the Most Weight



Two or more prior sentencings contributes 8 points to General Reconviction and 0 to Violent Reconviction, while severe interference from mental health problems contributes 9 points to Violent Reconviction and 3 to General Reconviction. These examples further illustrate how the same information can have different predictive value depending on the outcome.

Although item weights help predict recidivism, they should not be interpreted as direct measures of treatment needs, which should continue to be identified through appropriate clinical and case-planning processes. The differences between the models also have an important practical implication. A single overall risk score can obscure variation in the type of outcome being predicted. Presenting the General and Violent Reconviction scores side by side allows teams to consider broad recidivism risk without assuming that it is interchangeable with violence-specific risk. A higher general score does not automatically imply an equivalently high violent score. Decisions should therefore refer to the specific outcome relevant to the question being considered.

Discussion: Implications for Alaska

The findings support a practical strategy for improving correctional assessment: using locally observed outcomes to reassess how existing information is combined. Alaska currently collects extensive information through the LSI-R and related records. Outcome-specific weighting makes it possible to extract a more accurate predictive model from those same data with limited additional collection burden. The value of this approach lies partly in preserving existing assessment workflows while changing how their information is weighted.

The separate models also clarify a distinction that is easy to lose in practice. General reconviction and violent reconviction are related but not interchangeable. A tool optimized for general criminal justice involvement may be insufficiently informative for decisions specifically concerned with violence. The observed differences in item contributions provide a rationale for maintaining two scores rather than treating violence risk as merely the upper end of one general risk continuum.

At the same time, the development results do not demonstrate that implementing the scorecard will improve treatment court outcomes. The models were evaluated retrospectively on historical probation cases. They have not yet established prospective accuracy among treatment court referrals, nor have they been evaluated as part of an admission or supervision decision process. The next stage is therefore not simply to select a numerical cutoff; it is to determine how the scores can improve the consistency and proportionality of decisions without displacing individualized clinical and judicial review.

Using the Scores and Next Steps

The localized LSI-R produces separate General and Violent Reconviction scores that can inform treatment court screening, supervision, and service decisions. However, the thresholds used to translate these scores into risk classifications have not yet been established. Although the statistical development process determines how individual factors contribute to predicted risk, it does not determine where agencies should draw the boundaries between risk levels. Those decisions depend on how the scores will be used, the populations being assessed, and the consequences of incorrectly classifying someone as higher or lower risk. A threshold appropriate for identifying individuals who may benefit from additional services, for example, may not be appropriate for determining eligibility for a treatment court.

Establishing these thresholds therefore requires Alaska stakeholders to first identify the decisions each score is intended to support. The two scores provide different but complementary information, and their practical value depends on how that information is incorporated into existing policies and procedures. Several potential applications illustrate how the scores could be used:

- **Treatment court screening:** The Violent Reconviction score should inform admission decisions by identifying candidates who may warrant additional review based on their

risk of violent reoffending. This would support more individualized decisions without relying exclusively on current charges, offense histories, or automatic exclusions based on a single score that combines general and violent reconviction risk.

- **Supervision classification:** The General Reconviction score should inform decisions about supervision intensity, consistent with the risk principle that higher-risk individuals generally benefit from more intensive interventions while intensive supervision of lower-risk individuals may be counterproductive. Classification thresholds would help Alaska agencies allocate supervision resources according to assessed risk.

- **Targeted intervention and services:** The Violent Reconviction score will help identify individuals who may benefit from additional assessment for violence-focused interventions, including domestic violence programming or cognitive-behavioral treatment. In this context, the score would serve as an initial indicator of risk rather than a substitute for evaluating individual treatment needs.

Considering the General and Violent Reconviction scores together allows agencies to distinguish between different types of risk when making supervision and treatment decisions. An individual with elevated general risk but relatively low violent risk may warrant more intensive supervision without necessarily requiring violence-focused intervention. Conversely, an individual with lower general risk but elevated violent risk may require a different combination of supervision, assessment, and services.

Although the project initially focused on developing a violence risk assessment for Alaska treatment court eligibility decisions, the resulting models have broader applications. Using information already collected for the state's probation population, the research team developed updated LSI-R models predicting both general and violent reconviction. Given their improved predictive performance relative to the existing LSI-R scoring, we recommend that Alaska consider adopting the updated models across its probation population to improve risk estimation while also supporting treatment court eligibility decisions.

Regarding next steps, we suggest the courts and probation agencies determine which populations and decisions the updated models will best be used and establish appropriate risk

classification thresholds. Alaska stakeholders should identify the proportion of individuals to be assigned to each risk category and the corresponding policy implications, including supervision intensity, treatment referrals, and treatment court eligibility. These priorities should guide the selection of score cutoffs that align with the intended uses of each model.

Implementation and Ongoing Evaluation

Implementation should specify when scores are calculated, who verifies the underlying OST and administrative inputs, how missing or disputed information is handled, and how the results of the general and violent models are presented to decisionmakers (All Rise, 2026). Written guidance should also address the documentation of overrides and the use of scores in treatment planning. Clear procedures can help to promote consistent calculation and interpretation across settings.

After implementation, Alaska should monitor score distributions, missingness, item coding, use of overrides, and the proportion of referrals assigned to each risk category. As three-year outcomes become available, the updated assessment should be revalidated, as indicated by industry standard guidelines (see Desmaris et al., 2022). Routine monitoring of risk scores and category proportions is also needed following implementation to ensure that the tool is functioning as intended

Recommended Next Steps

Implementing the updated LSI-R will require Alaska to establish risk classification thresholds, integrate the scoring procedures into existing workflows, and prepare staff to use the results consistently. Therefore, we offer four recommendations. First, Alaska stakeholders should determine how the General and Violent Reconviction scores will be used and establish risk classification thresholds that reflect the desired distribution of individuals across risk categories

and the associated implications for supervision, treatment, and eligibility decisions. Second, the updated scoring procedures should be integrated into Alaska's existing assessment workflow or case management system, allowing officers to calculate both scores without collecting additional information. Third, staff should receive training on how to calculate and interpret the scores, with particular emphasis on their distinct purposes and appropriate uses.

Fourth, following implementation, Alaska should routinely monitor risk classifications and compare predicted probabilities with observed reconviction rates. Once sufficient follow-up data have accumulated, the updated LSI-R should be revalidated using newly assessed individuals to determine whether its predictive performance remains consistent over time. Together, these recommendations provide a practical framework for implementing the updated LSI-R while supporting its continued accuracy and usefulness in correctional decision-making.

Limitations

While the localized LSI-R was developed from Alaska's own population and evaluated on individuals who played no role in its development, it is not without limitations. The following limitations identify where the findings should be interpreted cautiously. They also describe areas requiring additional evidence before widespread operational use.

Development population and intended use. The assessment was developed using Alaska's probation-supervised population rather than treatment court participants specifically. Treatment court participants are selected for substance use or mental health needs and receive more intensive supervision and services, which may alter the relationships between assessment items and reconviction. Performance among treatment court participants should be examined once enough participants have been assessed and followed.

Conviction-based outcomes. Both outcomes rely on new convictions. Convictions capture only behavior that is detected, charged, and adjudicated, and plea agreements can result in conviction offenses that differ from the underlying behavior. Violent reconviction in particular may undercount violent behavior, and the assessment predicts officially recorded violence rather than all violence.

Race and ethnicity. Combining multiple racial and ethnic populations into a single Non-White category made stable evaluation possible, but it can conceal differences among the populations within that category, including Alaska Native individuals. Performance across racial and ethnic groups should continue to be evaluated as additional cases and outcomes accumulate following implementation. Future monitoring should prioritize more detailed subgroup estimates when the number of outcome events supports reliable comparisons. Until then, the combined category should be interpreted cautiously.

Informing rather than determining decisions. The assessment estimates reconviction risk using the information available in Alaska's records and does not capture every consideration relevant to an individual case. It is intended to provide a consistent, empirically derived source of information for treatment court teams and supervision staff, not to replace their judgment. A score should therefore support, rather than determine, the final decision.

These limitations underscore the importance of treating validation as an ongoing process rather than a one-time exercise. By building on Alaska's existing assessment infrastructure and using local outcome data to refine risk estimates, the project provides a stronger empirical foundation for treatment court screening and supervision decisions. Continued monitoring and collaboration with Alaska stakeholders can further improve the models' accuracy, fairness, and practical utility over time.

Conclusion

Alaska adopted the LSI-R, an instrument developed using Canadian samples nearly three decades ago, to bring structured assessment into probation practice. While the current tool provides a foundation for assessing risk, instruments developed elsewhere require local validation and periodic revalidation as populations, supervision practices, and intended uses evolve. The initial validation found that the LSI-R provided weak-to-moderate accuracy and substantially overstated the likelihood of reconviction, particularly for felony and violent outcomes. The validation results demonstrate why localization is not simply a matter of adopting a widely used assessment. Assessment must be periodically reevaluated using recent data and refined to support the decisions it is intended to inform.

The localized LSI-R demonstrates the potential value of deriving outcome-specific risk scores from information Alaska already collects. Using Alaska probation data, both the General and Violent Reconviction models achieved strong predictive discrimination, each identifying different combinations of factors relevant to each outcome. The improvement in predictive accuracy for both outcomes supports continued development of a two-score approach rather than reliance on a single general risk measure for decisions involving violence. Because the existing LSI-R substantially overstated violent reconviction, the calibration of the localized Violent Reconviction score is particularly relevant to decisions informed by predicted probabilities.

Moving forward, Alaska stakeholders should define how the assessments will be used in practice, establish decision thresholds and their implications for case management, and continue collecting assessment and outcome data to support future validation and refinement. With appropriate implementation and continued validation, the localized scores could provide a more consistent empirical basis for screening and supervision decisions while preserving the role of

professional judgment and individualized treatment planning. Ultimately, the value of the updated LSI-R lies not simply in more accurate risk prediction, but in helping Alaska make better-informed supervision and treatment decisions.

References

All Rise. (2026). *Adult treatment court best practice standards* (2nd ed., updated January 2026). https://allrise.org/publications/standards/

Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory-Revised*. Multi-Health Systems.

Andrews, D. A., Bonta, J., & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice and Behavior, 17*(1), 19–52.

Bonta, J., & Andrews, D. A. (2007). *Risk-need-responsivity model for offender assessment and rehabilitation*. Public Safety Canada.

Bucklen, K. B., Duwe, G., & Taxman, F. S. (2021). *Guidelines for post-sentencing risk assessment*. National Institute of Justice.

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36*(6), 567–590.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*(3), 837–845.

Desmarais, S. L., D'Amora, D. A., & Tavárez, L. P. (2022). *Advancing fairness and transparency: National guidelines for post-conviction risk and needs assessment*. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Assistance.

Duwe, G. (2014). The development, validity, and reliability of the Minnesota Screening Tool Assessing Recidivism Risk (MnSTARR). *Criminal Justice Policy Review, 25*(5), 579–613.

Duwe, G. (2024). Evaluating bias, shrinkage and the home-field advantage: Results from a revalidation of the MnSTARR 2.0. *Corrections, 9*(1), 20–42.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34*(4), 575–608.

Hamilton, Z., Kigerl, A., & Kowalski, M. (2022). Prediction is local: The benefits of risk assessment optimization. *Justice Quarterly, 39*(4), 722–744.

Hamilton, Z., Kigerl, A., Allen, B., Ursino, J., & Krushas, A. (2025). Never going to let you down: Preventing predictive shrinkage via the STRONG-R assessment method. *Justice Quarterly, 42*(7), 1279–1299. https://doi.org/10.1080/07418825.2024.2386637

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime & Delinquency, 52*(1), 77–93.

National Association of Drug Court Professionals. (2018). *Adult drug court best practice standards: Volume I* (Rev. ed.).

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2014). Thirty years of research on the Level of Service scales: A meta-analytic examination of predictive accuracy and sources of variability. *Psychological Assessment, 26*(1), 156–176.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's *d*, and *r*. *Law and Human Behavior, 29*(5), 615–620.

Steyerberg, E. W. (2019). *Clinical prediction models: A practical approach to development, validation, and updating* (2nd ed.). Springer. https://doi.org/10.1007/978-3-030-16399-0

Ursino, J., & Hamilton, Z. (2026). *Validating the Alaska Level of Service Inventory-Revised (LSI-R) – Research brief* (Doc. #2026-003). Advancing Research in Corrections Lab, University of Nebraska Omaha.

Vose, B., Cullen, F. T., & Smith, P. (2008). The empirical status of the Level of Service Inventory. *Federal Probation, 72*(3), 22–29.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32–35.

Technical Appendix

This appendix documents the statistical specifications and development decisions needed to understand and reproduce the localized LSI-R. The main report describes these procedures in practitioner terms. The specifications should be read alongside the substantive interpretation in the main report.

A1. Statistical Modeling Specifications

Coding and data preparation. Each candidate item was coded so that higher values indicate greater risk. Dichotomous LSI-R items were coded 1 when the risk factor was present and 0 otherwise. The thirteen rated items were dichotomized following standard LSI-R scoring, with original ratings of 0 or 1 coded 1 and ratings of 2 or 3 coded 0. Age was grouped into six ordered categories coded 0 (61 or older) through 5 (20 or younger) and entered as a single ordinal term so that younger age corresponds to greater risk. Counts of prior felony, violent, and total offenses were entered as counts. Cases missing the outcome were excluded, and missing predictor values were replaced with the item median. Race was retained for evaluation only and excluded from model construction.

Estimation and tuning. Each model was estimated as a ridge (L2 penalized, $\alpha = 0$) logistic regression (Hoerl & Kennard, 1970). The penalty parameter λ was selected by five-fold cross-validation within the training sample, using binomial deviance as the criterion. The value minimizing mean cross-validated deviance (λ.min) was selected.

Models for both outcomes were computed using a penalized ridge logistic regression, a statistical approach that estimates how strongly each assessment item is associated with reconviction while reducing the risk of assigning too much importance to individual factors.

Unlike the LSI-R's original scoring structure, which assigns one point to each item, this approach uses observed reconviction outcomes to determine how much each factor should contribute to the overall risk estimate. Ridge regression also addresses a common challenge in risk assessments, where many items capture related characteristics and therefore provide overlapping information. Without accounting for this overlap, a model may assign unusually large weights to certain items or produce weights that change substantially across samples. Ridge regression reduces this problem by applying a penalty that pulls item weights toward zero, limiting the influence of less informative factors while allowing stronger predictors to retain greater influence (Hoerl & Kennard, 1970; Friedman et al., 2010). Importantly, all candidate items remain in the model, but their contributions are adjusted to produce more stable predictions.

The appropriate strength of this penalty was determined using a process called five-fold cross-validation, which evaluates how well a model performs on data that were not used to estimate its weights. The training sample was divided into five groups. The model was developed using four groups and then evaluated on the remaining group to assess how well it predicted reconviction among individuals whose records were not used to estimate that version of the model. This process was repeated five times, with each group serving as the evaluation group once. Different penalty strengths were compared, and the one producing the best average predictive performance across the five evaluations was selected. This approach helps identify a model that performs consistently across different portions of the data rather than one that fits the training records particularly well but performs poorly on new cases. The separate test sample was then reserved for the final evaluation of the selected model.

Direction constraint. Coefficients from the initial model were inspected, and predictors with negative coefficients were removed. The model was re-estimated with the retained

predictors using the same cross-validation procedure, and λ.min was again selected. This procedure was completed separately for the General and Violent Reconviction models.

Evaluation. Final models were applied to the test sample to generate predicted probabilities. AUC was estimated with 95% confidence intervals using the DeLong method (DeLong et al., 1988). Calibration was assessed by grouping predicted probabilities into deciles and comparing mean predicted probability with the observed rate in each decile (Steyerberg, 2019). The ECE was calculated as the mean absolute difference between predicted and observed rates across deciles, weighted by decile size, with a 95% confidence interval from 1,000 bootstrap resamples of the test sample. Classification thresholds for subgroup error rates were identified using Youden's J (Youden, 1950).

A2. Items Without Weight in Either Model

Eleven LSI-R items received no weight in either final model and do not appear on the scorecard: any prior adult convictions (item 1), charge laid or supervision suspended during prior community supervision (item 9), suspended or expelled at least once (item 17), financial problems (item 21), dissatisfaction with marital or equivalent situation (item 23), non-rewarding parental relationships (item 24), non-rewarding relationships with other relatives (item 25), few anti-criminal acquaintances (item 35), few anti-criminal friends (item 36), substance use interfering with school or work (item 43), and poor attitude toward supervision (item 54). Information on prior convictions remains represented through items 2 and 3 and the counts of prior offenses. Items scored zero for only one outcome are shown in Table A1. Exclusion from the operational scorecard means these items did not provide sufficient additional predictive information once the other factors in the models were taken into account.

A3. Complete Point Values

Table A1. Point Values by Item and Outcome

Subcomponent	Item	Response Scored	General	Violent
Age and Prior Record	Age	20 or younger	11	13
Age and Prior Record	Age	21 to 30	8	10
Age and Prior Record	Age	31 to 40	6	8
Age and Prior Record	Age	41 to 50	4	5
Age and Prior Record	Age	51 to 60	2	3
Age and Prior Record	Age	61 or older	0	0
Age and Prior Record	Prior felony offenses	Per offense	0	0
Age and Prior Record	Prior violent offenses	Per offense	1	2
Age and Prior Record	Total prior offenses	Per offense	0	0
Age and Prior Record	Sex	Male	0	0
Criminal History	2. Two or more prior sentencings	Yes	8	0
Criminal History	3. Three or more prior sentencings	Yes	4	4
Criminal History	4. Three or more present offenses	Yes	2	0
Criminal History	5. Arrested under age 16	Yes	6	3
Criminal History	6. Ever incarcerated upon conviction	Yes	3	4
Criminal History	7. Escape history from an institution	Yes	2	0
Criminal History	8. Ever punished for institutional misconduct	Yes	0	1
Criminal History	10. Official record of assault or violence	Yes	3	0
Education / Employment	11. Currently unemployed	Yes	1	1
Education / Employment	12. Frequently unemployed	Yes	0	0
Education / Employment	13. Never employed for a full year	Yes	0	0

Subcomponent	Item	Response Scored	General	Violent
Education / Employment	14. Ever fired	Yes	4	1
Education / Employment	15. Less than regular grade 10	Yes	0	2
Education / Employment	16. Less than regular grade 12	Yes	0	0
Education / Employment	18. Participation / performance	Rating of 0 or 1	2	2
Education / Employment	19. Peer interactions	Rating of 0 or 1	1	1
Education / Employment	20. Authority interactions	Rating of 0 or 1	3	1
Financial	22. Reliance upon social assistance	Yes	0	1
Family / Marital	26. Criminal family or spouse	Yes	1	3
Accommodation	27. Unsatisfactory accommodation	Rating of 0 or 1	3	3
Accommodation	28. Three or more address changes in last year	Yes	1	1
Accommodation	29. High crime neighborhood	Yes	1	0
Leisure / Recreation	30. Absence of recent participation in organized activities	Yes	2	4
Leisure / Recreation	31. Could make better use of time	Rating of 0 or 1	0	1
Companions	32. Social isolate	Yes	9	12
Companions	33. Some criminal acquaintances	Yes	0	2
Companions	34. Some criminal friends	Yes	0	0
Alcohol / Drug Problems	37. Alcohol problem, ever	Yes	1	0
Alcohol / Drug Problems	38. Drug problem, ever	Yes	3	2
Alcohol / Drug Problems	39. Alcohol problem, currently	Rating of 0 or 1	2	1
Alcohol / Drug Problems	40. Drug problem, currently	Rating of 0 or 1	2	0
Alcohol / Drug Problems	41. Law violations	Yes	6	2
Alcohol / Drug Problems	42. Marital / family problems	Yes	2	0

Subcomponent	Item	Response Scored	General	Violent
Alcohol / Drug Problems	44. Medical problems	Yes	10	6
Alcohol / Drug Problems	45. Other clinical indicators	Yes	0	3
Emotional / Personal	46. Moderate interference	Yes	2	1
Emotional / Personal	47. Severe interference, active psychosis	Yes	3	9
Emotional / Personal	48. Mental health treatment, past	Yes	6	3
Emotional / Personal	49. Mental health treatment, present	Yes	1	3
Emotional / Personal	50. Psychological assessment indicated	Yes	2	2
Attitudes / Orientation	51. Supportive of crime	Rating of 0 or 1	0	1
Attitudes / Orientation	52. Unfavorable toward convention	Rating of 0 or 1	0	0
Attitudes / Orientation	53. Poor attitude toward sentence	Yes	1	0

Responses not listed, including the absence of each risk factor and ratings of 2 or 3 on rated items, score zero. Items showing zero points for both outcomes were retained by at least one model but carry weights that round to zero.

A4. Classification Error Rates at Alternative Thresholds

To provide a starting point for examining classification errors, initial thresholds were identified using Youden's J, a statistical method that balances the ability to identify reconvictions against correctly identifying individuals who are not reconvicted (Youden, 1950). This produced reference cutoffs of 33% predicted probability for General Reconviction and 11% for Violent Reconviction. These thresholds are not recommended classification or eligibility cutoffs. Rather, they provide a consistent reference point for examining how classification errors vary across groups.

Table A2 presents the Violent Reconviction Youden threshold alongside three more selective alternatives in the test sample. It reports the percentage flagged, the percentage

reconvicted among those flagged, and false-positive and false-negative rates for the full sample and separately for men and women. No single cutoff is automatically appropriate for Alaska's agencies. These figures are intended to support local discussion of the consequences of alternative thresholds, rather than establish an operational classification rule.

Table A2. Violent Reconviction Classification Error Rates at Alternative Thresholds, Test Sample

Threshold	Group	Flagged (%)	Reconvicted Among Flagged (%)	False Positive Rate (%)	False Negative Rate (%)
Youden (0.11)	All	49.0	28.8	41.8	14.3
	Men	49.3	27.4	42.5	14.7
	Women	48.2	32.5	39.7	13.3
Top 20% (0.28)	All	20.1	40.0	14.5	51.0
	Men	21.4	34.8	16.6	52.9
	Women	16.9	57.1	8.8	46.7
Probability of 0.10 or higher	All	51.7	27.3	45.0	14.3
	Men	52.1	25.9	45.9	14.7
	Women	50.6	31.0	42.6	13.3
Top 10% (0.36)	All	10.1	50.0	6.0	69.4
	Men	9.8	47.6	6.1	70.6
	Women	10.8	55.6	5.9	66.7

Note. Thresholds were applied to the Violent Reconviction model's predicted probabilities in the test sample (N = 586; base rate 16.4%). Values in parentheses are the predicted-probability cutoffs. The top 20% and top 10% thresholds classify the highest-scoring 20% and 10% of the full test sample. Flagged is the share of each group classified above the threshold.